用于独立特征学习的稀疏非负矩阵分解算法 *

黄卫春 [a]，赵　杨 [a+]，熊李艳 [b]

(华东交通大学 a. 软件学院; b. 信息工程学院, 南昌 330013)

摘　要：在保持非负矩阵分解（NMF）可解释性的同时，提高潜在特征的语义独立性是一个开放的研究问题。为了防止特征的共适应性，提出利用余弦相似度来减少潜在特征之间的相关性，从而提高 NMF 的独立特征学习能力。此外，为了使得分解后的矩阵具有较好的稀疏性，提出在传统 NMF 模型中引入 $L_{2,1/2}$ 稀疏约束，增强了算法的局部学习能力和稳健性。因此，潜在特征中的语义信息更加明显，潜在空间的表示更具有判别性。在 fetch_20newsgroups 数据集上对文档聚类的实验结果表明，提出的 INMF 算法在一系列评价指标上效果都优于传统的 NMF、SNMF 等算法模型。

关键词：非负矩阵分解；$L_{2,1/2}$ 稀疏；独立特征学习；余弦相似性

中图分类号：TP391.1　　doi: 10.19734/j.issn.1001-3695.2018.10.0724

Sparse non-negative matrix factorization algorithm for independent feature learning

Huang Weichun[a], Zhao Yang[a+], Xiong Liyan[b]

(*a. Software College, Nanchang, b. Information Engineering College, East China Jiaotong University, Nanchang* 330013, *China*)

Abstract: It's an open research problem to improve the semantic independence of potential features while maintaining the interpretability of non-negative matrix decomposition (NMF) . To prevent the co-adaptability of features, cosine similarity is proposed to reduce the correlation between potential features so as to improve the independent feature learning ability of NMF. In addition, in order to make the decomposed matrix have better sparsity, $L_{2,1/2}$ sparse constraint is introduced into the traditional NMF model to enhance the local learning ability and robustness of the algorithm. Therefore, the semantic information in the latent features is more obvious, and the representation of the potential space is more discriminative. The experimental results of document clustering on fetch_20newsgroups dataset show that the proposed INMF algorithm is superior to the traditional NMF and SNMF algorithm models in a series of evaluation indexes.

Key words: non-negative matrix factorization; $L_{2,1/2}$ sparse; independent feature learning; cosine similarity

0　引言

NMF 的目的是将原始高维数据矩阵分解为两个低维数据矩阵，并且两个低维数据矩阵的乘积尽可能地近似原始高维数据矩阵。

设一个原始的数据矩阵 $X=[x_1,x_2\cdots x_N]\in R_+^{M\times N}$ 。NMF 寻求将 X 分解成非负基矩阵 $U=[u_1,u_2\cdots u_K]\in R_+^{M\times K}$ 和非负系数矩阵 $V=[v_1,v_2\cdots v_N]\in R_+^{K\times N}$ ，其中 K 是数字潜在的特征。这也可以写成等效的矢量公式 $x_j\approx\sum_{i=1}^{k}u_iV_{ij}$ 。通常情况下，有 $K<<\min(M,N)$ 用来降维，v_j 是 U 列上的原始数据向量 x_j 的权重系数。NMF 将数据矩阵分解为基本的线性组合向量。

很多研究学者尝试采用不同的代价函数来衡量 NMF 的性能，他们的研究重点都是找到使得该损失函数可以取得最小值的因子矩阵 (U,V) 。目前，很多研究人员已经提出了一些方法来改进 NMF，如基于 Fisher 判别分析的增量式非负矩阵分解算法[1]、基于 Hessian 正则化的多视图联合非负矩阵分解算法[2]、稀疏约束图正则非负矩阵分解的增量学习算法[3]等。韩素青等人[4]提出一种通过在非负矩阵分解过程中对基矩阵列向量施加 L_1 与 L_2 范数稀疏约束,首先挖掘嵌入在高维数据中的低维数据结构,实现高维数据的低维表示；然后利用在低维数据聚类中性能良好的 K-Means 算法对稀疏降维后的数据进行聚类；孙静等人[5]提出一种基于特征融合的多约束非负矩阵分解算法。该算法不仅考虑了少量已知样本的标签信息和稀疏约束，还对其进行了图正则化处理，而且将分解后的具有不同稀疏度的图像特征进行了融合，从而增强了算法的聚类性能和有效性。

然而潜在特征之间的先天相关性使得传统的 NMF 算法缺乏辨别力，并且增加了最小化损失函数的困难。此外，传统的 NMF 算法没有充分利用矩阵的稀疏性，并且忽略了不同特征之间的相关性的有用信息。

稀疏特征选择旨在应用各种稀疏模型来实现特征选择并实现数据的稀疏表示。许多研究[15~17]已经将 L_1 范数扩展到 L_P 范数 $(0<p<1)$ 以获得更好的稀疏性。在文献[18,19]中，徐等人已经得出结论，当 p 为 1/2 时，L_P 范数，即 $L_{1/2}$ 范数具有最好的稀疏性。在文献[20]中，Nie 等人已经引入了联合 $L_{2,1}$ 范数最小化的损失函数和正则化的特征选择。然而 $L_{2,1}$ 范数不具备很好的稀疏性，因为它是基于 L_1 范数。最近，Wang 等人[21]

收稿日期：2018-10-04；**修回日期：**2018-12-05　**基金项目：**国家自然科学基金资助项目（61562027）；江西省科技厅资助项目（201616BBI90032，20161BAB212050）

作者简介：黄卫春（1968-），男，江西临川人，教授，硕导，硕士，主要研究方向为高性能计算、数据挖掘；赵杨（1994-），男（通信作者），湖北十堰人，硕士研究生，主要研究方向为数据挖掘（991470264@qq.com）；熊李艳（1969-），女，江西南昌人，教授，硕导，主要研究方向为软件工程、数据挖掘.

提出了将 $L_{2,1}$ 范数扩展到 $L_{2,p}$ 矩阵范数 $(0<p<1)$ 的思想，以便选择联合的、更稀疏的特征，同时该模型比 $L_{2,1}$ 范数具有更好的鲁棒性。实验证明，当 p 等于 1/2 时，$L_{2,p}$ 矩阵范数具有最好的性能，因此本文将 $L_{2,1/2}$ 矩阵范数模型应用于的新 NMF 以实现稀疏约束。

在本文中提出利用余弦相似度来减少潜在特征之间的相关性，从而提高 NMF 的独立特征学习能力。此外，本文引入了 $L_{2,1/2}$ 稀疏约束到 INMF 中。因此，潜在特征中的语义信息更加清晰，潜在空间中的表示更具判别力。本文将这些方法与文档聚类中的几种基本 NMF 模型进行了比较，并给出了本文算法和传统算法在真实数据集上的实验结果对比。本文工作的主要贡献可以总结如下：

a)本文提出的 INMF 通过引入余弦相似性防止特征共适应，从而改进 NMF；

b)在特征空间的基本矩阵上添加 $L_{2,1/2}$ 范数的稀疏约束，不仅可以实现数据的稀疏表示，简化计算，还可以提高算法的局部学习能力和鲁棒性。

1　相关工作

人们从不同的角度对 NMF 进行了改进研究，如约束 NMF[6~9]、结构化 NMF[10.11]和广义 NMF[12~14]。最常见的约束 NMF 是稀疏 NMF，稀疏性约束有助于改进分解的唯一性，以及得到基于局部的表示，通常通过 L_1 范数[6]来实现特征选择和数据的稀疏表示。正交 NMF 具有良好的性能，因为正交 NMF 的结果对应于一个独特的解区域中的稀疏区域，它学习得最明显[7]。图正则化 NMF（GRNMF）提高了文档和图像等聚类任务中的性能，它通过在散布的数据点上构建最近的邻域图来模拟流形结构[8,9]。公式加权是学习算法的常用改进方法，可以用来强调不同成分的相对重要性。加权 NMF 在协同过滤和聚类任务中很流行，因为它们根据实例的连接将先验知识合并到损失函数中[10,11]。对于具有复杂和异构信息源的任务，提出了广义 NMF，如 Semi-NMF[12]、非负张量分解[13]和非负矩阵集因式分解[14]。

一种特殊的方法 dropout NMF[22]可以通过改变潜在特征的更新过程来防止隐藏单元的相互适应。由于固定的共同作用被打破，隐藏的单元仍然可以从别人那里学习，但依赖性较小。受这些启发，本文提出了一种新的 NMF，其通过最小化潜在特征之间的余弦相似性来起作用。本文发现通过打破潜在特征之间的相关性可以改进 NMF。因此，在下一章中，将余弦相似性纳入 NMF。此外，添加在系数矩阵 U 上的 $L_{2,1/2}$ 稀疏约束用于选择最具判别性的稀疏特征。

2　方法

NMF 的目的是让系数矩阵 U 和基本矩阵 V 的乘积尽可能接近原始数据矩阵 X。NMF 的公式如下：

$$X \approx UV^T \tag{1}$$

应该最小化平方欧几里德距离损失函数：

$$L=\|X-UV^T\|_F^2,\ s.t. U\in R_+^{M\times K},\ V\in R_+^{K\times N} \tag{2}$$

在本文中采用平方欧几里德距离损失函数：

$$L=\|X-UV^T\|_2^2=tr(X^TX)-2tr(X^TUV)+tr(V^TU^TUV) \tag{3}$$

其中：$\|\cdot\|_2^2$ 是 L_2 范数的平方；$tr(\cdot)$ 代表矩阵的迹。梯度下降算法的迭代更新规则如下：

$$u_{mk}\leftarrow u_{mk}\frac{(XV^T)_{mk}}{(UVV^T)_{mk}},\leftarrow v_{kn}\frac{(U^TX)_{kn}}{(U^TUV)_{kn}} \tag{4}$$

2.1　用于独立特征学习的稀疏非负矩阵分解

NMF 可以表示为线性神经网络，因为输入 x_i 由 U 中的基向量的线性组合表示：

$$x_i=Uv_i=\sum_k v_{ki}u_k \tag{5}$$

其中：u_k 是第 k 个潜在特征。

由于潜在特征在 NMF 中是相关的，共适应是指 $\{u_k\}_{k=1}^k$ 的更新停止在鞍点处的状态，其中 L 仍然可以进一步优化，直到达到最大迭代。研究人员试图在同一数据集上通过在不同的初始化策略下执行改进的 NMF 来避免共适应，这带来了很大的计算复杂性[18]。 因此，本文提出了一种新的 NMF，在每次迭代中，通过最小化潜在特征之间的余弦相似度，u_k 将被独立更新，潜在特征将逐渐不相关。

2.1.1 稀疏约束

虽然 NMF 可以降低原始数据的维数，但如何选择判别特征和实现数据的稀疏表示仍然是复杂的。因此，$L_{2,1/2}$ 范数的疏约束被引入作为附加条件添加在基本矩阵 U 上。它可以写成

$$\|U\|_{2,1/2}=\left(\sum_{i=1}^m\|U_i\|_2^{1/2}\right)^2 \tag{6}$$

2.1.2 余弦相似性度量

Theodoridis 等人[24]将余弦相似性度量定义为 $\text{Scosine}(x,y)=\frac{x^Ty}{\|x\|\|y\|}$，其中：$\|x\|=\sqrt{\sum_{i=1}^l x_i^2}$ 和 $\|y\|=\sqrt{\sum_{i=1}^l y_i^2}$ 分别是矢量 x 和 y 的长度；x 和 y 都是 l 维向量。由于余弦相似度容易解释且对于稀疏向量计算简单，所以其被广泛用于文本挖掘和信息检索[25]。

2.1.3 迭代更新规则

考虑到上述因素，本文将独立特征学习的稀疏非负矩阵分解的目标函数定义如下：

$$L=\|X-UV\|_F^2+\theta\|U\|_{2,1/2}^{1/2}+\alpha\|V\|_2+\beta\sum_{i,j}\cos(U_i,U_j) \tag{7}$$

其中：θ、α、β 是非负的,它们可以平衡后面几个重建误差项的权重,θ、α 是稀疏参数,β 是余弦相似性参数。

为了优化这个目标函数，可以将式（7）中的目标函数转换如下：

$$\begin{aligned}L=&Tr(X^TX)-2Tr(X^TUV)+Tr(V^TU^TUV)+\\&\alpha(Tr(V^TV))+4\theta Tr(U^TQU)+\beta Tr(U^TUS)\end{aligned} \tag{8}$$

$Q=\left[q_{ij}\right]\in R^{m\times m}$ 是一个对角矩阵,可以如下计算对角矩阵 Q 的第 i 个对角元素 q_{ij}：

$$q_{ij}=\frac{1}{4\|U_i\|_2^{3/2}} \tag{9}$$

为了避免溢出，在定义矩阵 Q 时添加一个足够小的常数 ε，因此式（8）也可以写成如下形式：

$$q_{ij}=\frac{1}{4\max\left(\|U_i\|_2^{\frac{3}{2}},\varepsilon\right)} \tag{10}$$

其中：$S=[s_{ij}]\in R^{k\times k}$。可以计算矩阵 S 的元素 S_{ij} 如下：

$$S_{ij}=\sum_{i,j}^k\cos(U_i,U_j) \tag{11}$$

为了获得基本矩阵 U 和系数矩阵 V 的迭代更新规则，取 L 的偏导数：

$$\frac{\partial L}{\partial U}=-2XV^T+2UVV^T+8\theta QU+2\beta US \tag{12}$$

$$\frac{\partial L}{\partial V}=-2U^TX+2U^TUV+2\alpha V \tag{13}$$

基本矩阵 U 和系数矩阵 V 的迭代更新规则如下所示：

$$U_{mk}\leftarrow U_{mk}\frac{(XV^T)_{mk}}{(UVV^T+4\theta QU+\beta US)_{mk}} \tag{14}$$

$$V_{kn}\leftarrow V_{kn}\frac{(U^TX)_{mk}}{(U^TUX+\alpha V)_{kn}} \tag{15}$$

2.2 收敛性分析

本节将分析算法的收敛性，证明式（7）中的目标函数在迭代更新规则式（14）∽（15）中单调递减。

首先分析迭代更新规则式（14）的收敛性。

引理 2.1 [26]

$$\sum_{i=1}^m\left(\left\|g_i^{t+1}\right\|_2^{1/2}-\frac{\left\|g_i^{t+1}\right\|_2^2}{4\left\|g_i^t\right\|_2^{3/2}}\right)\le\sum_{i=1}^m\left(\left\|g_i^t\right\|_2^{1/2}-\frac{\left\|g_i^t\right\|_2^2}{4\left\|g_i^t\right\|_2^{3/2}}\right) \tag{16}$$

证明　从引理 2.1 有

$$\sum_{i=1}^m\left(\left\|U_i^{t+1}\right\|_2^{1/2}-\frac{\left\|U_i^{t+1}\right\|_2^2}{4\left\|U_i^t\right\|_2^{3/2}}\right)\le\sum_{i=1}^m\left(\left\|U_i^t\right\|_2^{1/2}-\frac{\left\|U_i^t\right\|_2^2}{4\left\|U_i^t\right\|_2^{3/2}}\right) \tag{17}$$

定义一个函数如下：

$$\begin{aligned}H(U,V)&=Tr(X^TX)-2Tr(X^TUV)+Tr(V^TU^TUV)\\&+\alpha(Tr(V^TV))+\beta Tr(U^TUS)\end{aligned} \tag{18}$$

因为 $\|U\|_{2,1/2}^{1/2}=\sum_{i=1}^m\|U_i\|_2^{1/2}$，可以得到以下不等式（19）：

$$H^{t+1}+\theta\sum_{i=1}^m\frac{\left\|U_i^{t+1}\right\|_2^2}{4\left\|U_i^t\right\|_2^{3/2}}=H^{t+1}+\theta\left\|U^{t+1}\right\|_{2,1/2}^{1/2}+\theta\sum_{i=1}^m\left(\frac{\left\|U_i^{t+1}\right\|_2^2}{4\left\|U_i^t\right\|_2^{3/2}}-\left\|U_i^{t+1}\right\|_2^{1/2}\right)\le$$

$$H^t+\theta\sum_{i=1}^m\frac{\left\|U_i^t\right\|_2^2}{4\left\|U_i^t\right\|_2^{3/2}}=H^t+\theta\left\|U^t\right\|_{2,1/2}^{1/2}+\theta\sum_{i=1}^m\left(\frac{\left\|U_i^t\right\|_2^2}{4\left\|U_i^t\right\|_2^{3/2}}-\left\|U_i^t\right\|_2^{1/2}\right) \tag{19}$$

结合式（15）和（18），可以得到以下不等式：

$$H^{t+1}+\theta\left\|U^{t+1}\right\|_{2,1/2}^{1/2}\le H^t+\theta\left\|U^t\right\|_{2,1/2}^{1/2} \tag{20}$$

因此，式（6）中的目标函数在迭代更新规则式（14）中单调减少。

接下来继续分析迭代更新规则式（15）的收敛性。

定义 2.1　$G(h,h')$ 是 $F(h)$ 的一个辅助函数如果满足以下条件：

$$G(h,h')\ge F(h),\ G(h,h)=F(h) \tag{21}$$

引理 2.2　如果 G 是一个辅助函数，那么 F 在以下更新下是非增的：

$$h^{t+1}=\arg\min_h G(h,h^t) \tag{22}$$

显然能够证明

$F(h^{t+1})\le G(h^{t+1},h^t)\le G(h^t,h^t)=F(h^t)$ 并且 $F(h)$ 是收敛的。

引理 2.3

$$G(V,V^t)=F(V^t)+F'(V^t)(V-V^t)+\frac{U^TUV+\alpha V}{V^t}(V-V^t)^2 \tag{23}$$

是 $F(V)$ 的辅助函数。

证明　$F(V)$ 求一阶导数和二阶导数分别是

$$F'(V)=\left(-2U^TX+2U^TUV+2\alpha V\right)_{kj},$$

$$F''(V)=\left(2U^TU+2\alpha\right)_{kk}$$

所以 $F(V)$ 的泰勒展开式可以写成如下形式：

$$F(V)=F(V^t)+F'(V^t)(V-V^t)+(U^tU+\alpha)(V-V^t)^2 \tag{24}$$

因为 $\left(U^TUV\right)_{kj}=\sum_h\left(U^TU\right)_{kh}V_{hj}^t\ge\left(U^TU\right)_{kk}V_{kj}^t$

$$\left(\alpha V\right)_{kj}=\alpha V_{kj}^t$$

有 $\dfrac{U^TUV+\alpha V}{V^t}\ge U^TU+\alpha$

所以 $G(V,V^t)\ge F(V)$

通过联立方程式（22）和（23），知道 $G(V^{t+1},V^t)$ 是式（23）的局部极小值，并且 V^{t+1} 是相应的局部最小点。

$G(V,V^t)$ 是 $F(V)$ 的辅助函数,因此 F 通过式(15)单调递减。

3　实验结果和分析

3.1　数据集

本文选择公开的语料库 fetch_20newsgroups 数据集进行实验。数据集大约收集了 20 000 份新闻文件的集合，可均匀地划分为 20 个不同的主题类别。它最初由朗肯收集，包含 18 846 个文件，并且在本文预处理之后仅保留 1 000 个词。每个文档被转换为向量 $x_t\in R^{1000}$，使得数据矩阵 $X\in R^{18846\times1000}$。本文将标准 NMF、稀疏 NMF（SNMF）和本文的新算法应用于这个 X，用于比较研究算法的准确度和召回率。

3.2　实验设置

评价指标：在本文中使用三种评价指标，即准确度（precision）、召回率（recall）和 F 值（F-measure）/（F1-score）来评估上述聚类算法的聚类性能。

公式如下：

$$Recall=\frac{n_{ij}}{n_j} \tag{25}$$

其中：n_{ij} 是已知类别 i 中的文本属于聚类 j 中的个数；n_j 是已知类别 i 中的文档数；n_j' 是聚类 j 中的文档数。

F 值同时考虑了上述的文档聚类精度 P 和召回率 r 来计算其得分，它是一种复合指标，其值越大，聚类效果越好。

$$F_1\text{-score}=\frac{2\times precision\times recall}{precision+recall} \tag{26}$$

参数设置：对于 fetch_20newsgroups 数据集，新的 NMF 的参数 α、β，从区域{0.01~0.1}中选择，并且 SNMF 的正则化项的平衡参数 α，与上述相同。为了验证不同主题数下算法的性能，K 分别设置为{5,10,15,20}。本文将 INMF 算法与传统的 NMF、SNMF 算法进行了对比实验，详细的实验对比结果在下节进行分析。

3.3　聚类结果

表 1~4 显示了两种基本方法与本文的方法在 fetch_20newsgroups 数据集上的聚类结果比较，对比不同 K 下准确率、召回率、F1-score 指标。三种方法中的最佳表现都是粗体显示的。它表明，无论 K 的取值为多少，本文提出的新方法均比传统方法表现得更好。它展示了本文提出的方法在防止潜在特征共适应的有效性。

表 1　k=5 数据集上的结果

Table 1　Results on k=5 datasets

算法	Precision	Recall	F1-score
NMF	0.199	0.198	0.198
SNMF	0.205	0.206	0.206
INMF	**0.223**	**0.222**	**0.222**

表 2　k=10 数据集上的结果

Table 2　Results on k=10 datasets

算法	Precision	Recall	F1-score
NMF	0.100	0.099	0.099
SNMF	0.102	0.101	0.101
INMF	**0.109**	**0.108**	**0.108**

表 3　k=15 数据集上的结果

Table 3　Results on k=15 datasets

算法	Precision	Recall	F1-score
NMF	0.064	0.064	0.064
SNMF	0.066	0.066	0.066
INMF	**0.089**	**0.088**	**0.088**

表 4　k=20 数据集上的结果

Table 4　Results on k=20 datasets

算法	Precision	Recall	F1-score
NMF	0.051	0.051	0.051
SNMF	0.050	0.049	0.050
INMF	**0.059**	**0.058**	**0.058**

图 1~3 直观地表明了随着 K 值的变化，本文提出的方法 INMF 与其他两种方法聚类结果的对比，分别从准确度、召回率和 F 值来进行评估。

如图 1 所示，当 k 分别取 5、10、15、20 时，INMF 在准确度方面与传统 NMF、SNMF 的对比。通过条形图的对比，可以更加直观地看到，在不同 K 值下，本文的 INMF 算法的准确度都是要优于其他两种算法的。

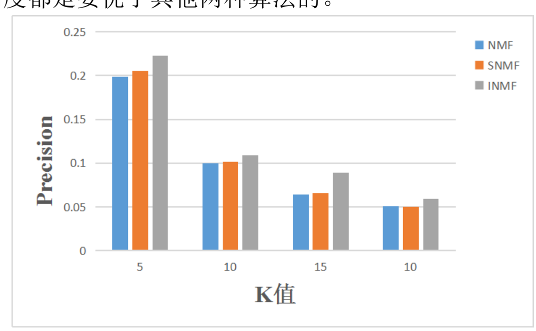


图 1　各 K 值下 NMF、SNMF、INMF 的准确性对比

Fig. 1　Accuracy comparison of NMF, SNMF and INMF under each K values

如图 2 所示，当 k 取不同值时，INMF 在召回率方面与传统 NMF、SNMF 的对比。通过条形图的对比，可以更加直观地看到，本文的 INMF 算法的召回率是要优于其他两种算法的，并且随着 K 值得增大优势更加明显。

如图 3 所示，当 k 取不同值时，INMF 在 F 值方面与传统 NMF、SNMF 的对比。通过条形图的对比，可以看到本文的 INMF 算法的 F 值是优于其他两种算法的，但总体表现随着 K 值得增大而降低，因为具有更多主题的更大的数据集对于聚类来说更困难。

3.4　参数选择与收敛性分析

当 p 分别设为{0.1，0.3，0.5，0.7，0.9}时，比较了在 k = 5 时，传统 NMF、SNMF 以及本文算法在该数据集上的收敛曲线。在图 4 中，显示了当 p 对应不同值时每种算法的结果。考虑到聚类性能和时间消耗，本文设置了最合适的参数。对比分析可以看到，所有算法在 20 次迭代内趋于收敛。在 fetch_20newsgroups 数据集上，NMF 和 SNMF 的迭代次数约为 18 和 19，目标函数趋于稳定。但 INMF 可以在 15 次迭代处收敛。这是因为通过本文作出的改进之后，每次迭代后更多潜在特征变得不相关的，累积的效果带来了更好的性能。

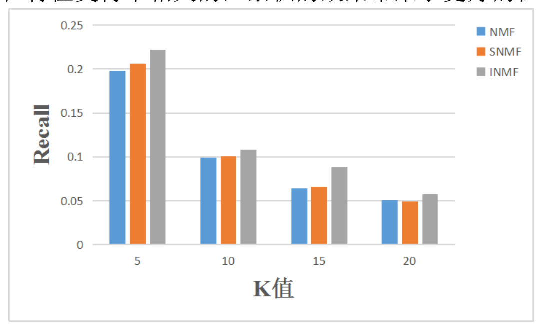


图 2　各 K 值下 NMF、SNMF、INMF 的召回率对比

Fig. 2　Comparison of recall rates of NMF, SNMF and INMF under each K values

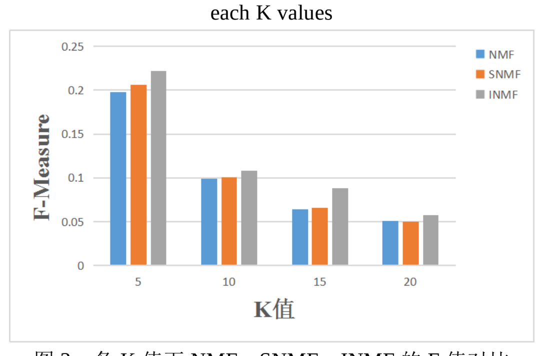


图 3　各 K 值下 NMF、SNMF、INMF 的 F 值对比

Fig. 3　Comparison of F-measure of NMF, SNMF and INMF under each K values

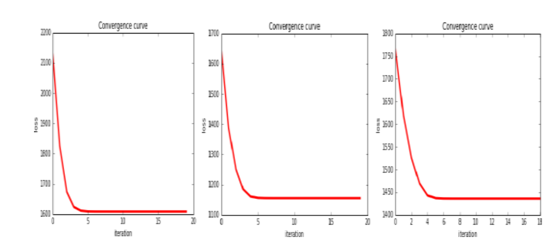


图 4　fetch_20newsgroups 上 NMF、SNMF 和 INMF 的收敛曲线

Fig. 4　Convergence Curves of NMF, SNMF and INMF on fetch_20 news groups

4　结束语

本文对传统 NMF 中潜在特征之间的相关性对算法性能的影响进行了分析，提出了独立特征学习的稀疏非负矩阵分解方法即 INMF。该方法不仅可以利用向量的余弦相似性，而且可以有效地学习目标的局部信息。在 INMF 中，潜在特征的更新是通过最小化每次迭代中的余弦相似度实现的。在本文所提出的算法中有效地防止了特征共适应性，因此潜在特征更加明确和更具有辨别力。此外，$L_{2,1/2}$ 稀疏约束作为 NMF 中的约束条件，可以使分解后的矩阵具有良好的稀疏性，增强了算法的局部学习能力和鲁棒性。从上面的实验结果中可以得出结论，本文提出的算法在精度和召回率方面都有较好的效果。未来，笔者将探索具有其他损失函数和约束条件的 NMF，并且将提出的新方法应用在不同的领域，如计算机视觉、语音识别、网络安全、生物医学工程等。

参考文献：

[1] 蔡竞，王万良，郑建炜，等. 基于 Fisher 判别分析的增量式非负矩阵分解算法 [J]. 模式识别与人工智能, 2018, 31(6): 505-515. (Cai Jing, Wang Wanliang, Zheng Jianbiao, *et al.* Incremental nonnegative matrix factorization algorithm based on fisher discriminant analysis [J]. Pattern Recognition & Artificial Intelligence, 2018, 31(6): 505-515.)

[2] 王超锋，施俊，吴金杰，等. 基于 Hessian 正则化的多视图联合非负矩阵分解算法 [J]. 计算机工程，2017, 43 (11): 134-139. (Wang Chaofeng, Shi Jun, Wu Jinjie, *et al.* Multi-view joint non-negative matrix factorization algorithm based on Hessian regularization [J]. Computer Engineering, 2017, 43 (11): 134-139.)

[3] 汪金涛，曹玉东，孙福明. 稀疏约束图正则非负矩阵分分解的增量学习算法 [J]. 计算机应用，2017, 37 (4): 1071-1074. (Wang Jintao, Cao Yudong, Sun Fuming. Incremental learning algorithm for regular non-negative matrix partition decomposition of sparse constrained graphs [J]. Journal of Computer Applications, 2017, 37 (4): 1071-1074.)

[4] 韩素青，贾茹. 基于稀疏约束非负矩阵分解的 K-Means 聚类算法 [J]. 数据采集与处理，2017, 32 (6): 1216-1222. (Han Suqing, Jia Ru. K-Means clustering algorithm based on sparse constrained nonnegative matrix factorization [J]. Journal of Data Acquisition & Processing, 2017, 32 (6): 1216-1222.)

[5] 孙静，蔡希彪，孙福明. 基于特征融合的多约束非负矩阵分解算法 [J]. 计算机应用, 2017, 37 (10): 2834-2840. (Sun Jing, Cai Xixi, Sun Fuming. Multi-constraint nonnegative matrix factorization algorithm based on feature fusion [J]. Journal of Computer Applications, 2017, 37 (10): 2834-2840.)

[6] Mohammadiha N, Leijon A. Nonnegative matrix factorization using projected gradient algorithms with sparseness constraints [C]// Proc of IEEE International Symposium on Signal Processing and Information Technology. New York: IEEE Conference Proceedings, 2009: 418-423.

[7] Ding Chris, Li Tao, Peng Wei, *et al.* Orthogonal nonnegative matrixt-factorizations for clustering [C]// Proc of ACM Sigkdd International Conference on Knowledge Discovery & Data Mining. USA:Philadelphia, PA, 2006.

[8] Cai Deng, He Xiaofei, Han Jiawei, *et al.* Graph regularized non-negative matrix factorization for data representation [J]. IEEE Trans Pattern Anal Mach Intell, 2011, 33 (8): 1548-1560.

[9] Zhi Ruicong, Flierl Markus, Ruan Qiuqi, *et al.* Facial expression recognition based on graph-preserving sparse non-negative matrix factorization [C]// Proc of IEEE International Conference on Image Processing. [S. l.]: IEEE Signal Processing Society, 2010: 3293-3296.

[10] Mao Yun, Saul L K. Modeling distances in large-scale networks by matrix factorization [C]// Proc of the 4th ACM SIGCOMM Conference on Internet Measurement. New York: ACM Press, 2004: 278-287.

[11] Kim Y D, Choi S. Weighted nonnegative matrix factorization [C]// Proc of IEEE International Conference on Acoustics, Speech and Signal Processing. Washington DC: IEEE Computer Society, 2009: 1541-1544.

[12] Ding C, Li Tao, Jordan M I. Convex and semi-nonnegative matrix factorizations [J]. IEEE Transa on Pattern Analysis & Machine Intelligence, 2010, 32 (1): 45-55.

[13] Mørup M, Hansen L K, Arnfred S M. Algorithms for sparse nonnegative Tucker decompositions [J]. Neural Computation, 2008, 20 (8): 2112-2131.

[14] Li Le, Zhang Yujin. Non-negative matrix-set factorization [C]// Proc of Fourth International Conference on Image and Graphics. Piscataway, NJ: IEEE Press, 2007: 564-569.

[15] Foucart S, Lai Minjun, The sparest solutions of underdetermined linear system by Lq-minimization for [J]. Applied & Computational Harmonic Analysis, 2009, 26 (3): 395-407.

[16] Chartrand R. Exact reconstruction of sparse signals via nonconvex minimization [J]. IEEE Signal Processing Letters, 2007, 14 (10): 707-710.

[17] Chartrand R. Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data [C]// Proc of IEEE International Conference on Symposium on Biomedical Imaging. 2009: 262-265.

[18] Xu Zongben, Chang Xiangyu, Xu Fengmin, *et al.* L1//2 regularization: a thresholding representation theory and a fast solver [J]. IEEE Trans on Neural Networks & Learning Systems, 2012, 23 (7): 1013-1027.

[19] ZongBen Xu, Hai Zhang, Yao Wang, *et al.* L1//2regularization [J]. Science in China Series F :Information Science , 2010, 53 (6): 1159-1169.

[20] Nie Feiping, Huang Heng, Cai Xiao, *et al.* Efficient and robust feature selection via joint L2, 1-norms minimization [C]// Proc of International Conference on Neural Information Processing Systems. [S.l.]:Curran Associates Inc, 2010: 1813-1821.

[21] Wang Liping, Chen Songcan. l2, p-matrix norm and its application in feature selection [EB/OL]. (2013).Available: http://arxiv. org/abs/1303. 3987.

[22] He Zhicheng, Liu Jie, Liu Caihua, *et al.* Dropout non-negative matrix factorization for independent feature learning [M]// Natural Language Understanding and Intelligent Applications. [S.l.] : Springer International Publishing, 2016: 201-212.

[23] Langville A N, Meyer C D, Albright R, *et al.* Algorithms, initializations, and convergence for the nonnegative matrix factorization [J]. Eprint Arxiv, 2014.

[24] Theodoridis S, Koutroumbas K. Pattern recognition (fourth edition) [M]. [S.l.] :Academic Press, 2008.

[25] Dhillon I S, Modha D S. Concept decompositions for large sparse text data using clustering [J]. Machine Learning, 2001, 42 (1-2): 143-175.

[26] Shi Caijuan, Ruan Qiuiqi, An Gaoyun, *et al.* Hessian semi-supervised sparse feature selection based on l2, 1//2-matrix norm [J]. IEEE Trans on Multimedia, 2014, 17 (1): 16-28.